Original Investigations

Comparison of "B" Readers' Interpretations of Chest Radiographs for Asbestos Related Changes[1]

Joseph N. Gitlin, DPH, Leroy L. Cook, BA, Otha W. Linton, MSJ, Elizabeth Garrett-Mayer, PhD

Rationale and Objectives. The purpose of this study was to determine if chest radiographic interpretations by physicians retained by attorneys representing persons alleging respiratory changes from occupational exposure to asbestos would be confirmed by independent consultant readers.

Materials and Methods. For 551 chest radiographs read as positive for lung changes by initial "B" readers retained by plaintiffs' attorneys, 492 matching interpretative reports were made available to the authors. Six consultants in chest radiology, also B readers, agreed to reinterpret the radiographs independently without knowledge of their provenance. The film source, patient name, and other identifiers on each film were masked. The International Labor Office 1980 Classification of Chest Radiographs (ILO 80) was used with forms designed by the US National Institute of Occupational Safety and Health to record the consultants' findings. The results were compared with initial readings for film quality, complete negativity, parenchymal abnormalities, small opacities profusion, and pleural abnormalities using chi-square tests and kappa statistics.

Results. Initial readers interpreted study radiographs as positive for parenchymal abnormalities (ILO small opacity profusion category of 1/0 or higher) in 95.9% of 492 cases. Six consultants classified the films as 1/0 or higher in 4.5% of 2,952 readings. Statistical tests of these and other comparable data from the study showed highly significant differences between the interpretations of the initial readers and the findings of the consultants.

Conclusion. The magnitude of the differences between the interpretations by initial readers and the six consultants is too great to be attributed to interobserver variability. There is no support in the literature on x-ray studies of workers exposed to asbestos and other mineral dusts for the high level of positive findings recorded by the initial readers in this report.

Key Words. Asbestosis; chest x-ray interpretation; ILO classification; disability compensation.

In 2000, the authors were requested by attorneys active in asbestos compensation litigation to develop an acceptable method of obtaining reliable interpretations of chest radiographs. The methods and results of a multiple reader trial conducted in response to their request are presented in this report. The study design was a comparison of six independent readings of chest radiographs by qualified consultant "B" readers with single readings of the same radiographs by one of several initial B readers selected by plaintiffs' counsel.

Chest radiographs have been used in public health programs for detection of tuberculosis and for legally mandated examinations of coal miners and other workers exposed to mineral dusts, including asbestos. Under current federal regulations, coal miners, uranium miners and millers, and workers with asbestos or asbestos-containing products who claim occupationally related respiratory disease or disability must support their claims with a posteroanterior (PA) chest radiograph. The findings of these

Acad Radiol 2004; 11:843–856

[1] From the Department of Radiology, Johns Hopkins Medical Institutions, 14208 Northwyn Drive, Silver Spring, MD 20904, PennyPoor Consultants, Shenandoah Junction, WV, International Society of Radiology, Bethesda, MD, Department of Oncology, Division of Biostatistics, Johns Hopkins School of Medicine, Baltimore, MD. Received June 10, 2003; revision received November 5, 2003; revision received February 2, 2004; revision received February 23, 2004; revision received March 8, 2004; revision received March 22, 2004; revision accepted April 12, 2004. **Address correspondence to** J.N.G. e-mail: jgitlin@jhmi.edu

doi:10.1016/j.acra.2004.04.012

radiographs are reported using a classification system (ILO 80) devised by the International Labor Office (1). The "B reader" certification offered by the Public Health Service's National Institute for Occupational Safety and Health (NIOSH) is specified as a credential for interpreters in regulations (2–4). It also is accepted widely by adjudicators and legal counsel as a qualification for expert witnesses. Approximately 700 radiologists and other physicians hold B reader certification.

In recent years, lawsuits on behalf of hundreds of thousands of workers alleging harm to their health from occupational exposures to asbestos particles have become a substantial burden on the courts and on companies that fabricated asbestos products or used asbestos in their products. Lawsuits seeking billions of dollars in compensation are pending in federal and state courts. More than 60 US companies have sought voluntary bankruptcy to deal with such claims (5,6).

In the 1960s, Selikoff et al. (7–9) demonstrated that exposure to respirable asbestos particles can result, decades later, in impairment of pulmonary function and, less frequently, in the development of mesothelioma, a form of lung cancer associated with asbestos inhalation. Many asbestos workers, over a lifetime, demonstrate lung changes that may well be attributable to smoking, sarcoidosis, tuberculosis, or other disease processes. However, not all persons exposed to asbestos suffer respiratory impairment. Pulmonary function tests should be part of any clinical evaluation, but the PA chest radiograph remains the most common and generally accepted method of finding and describing pulmonary changes. The American Thoracic Society called the chest x-ray the most valuable examination in the clinical diagnosis of asbestosis (10).

A small number of B readers has made reputations with attorneys by consistently interpreting chest radiographs of asbestos claimants as positive in 90–100% of cases (11). Positivity in classifying a chest radiograph for compensation means noting a parenchymal abnormality and indicating a small opacity profusion category of 1/0 or higher in the ILO 80 system. The ILO system has been adopted in the United States and most other developed countries as a basis for collecting statistics on a country's experience with mineral dust workers and their work-related respiratory problems. But the significant and contentious use of the system in the United States has been in worker's compensation programs and litigation.

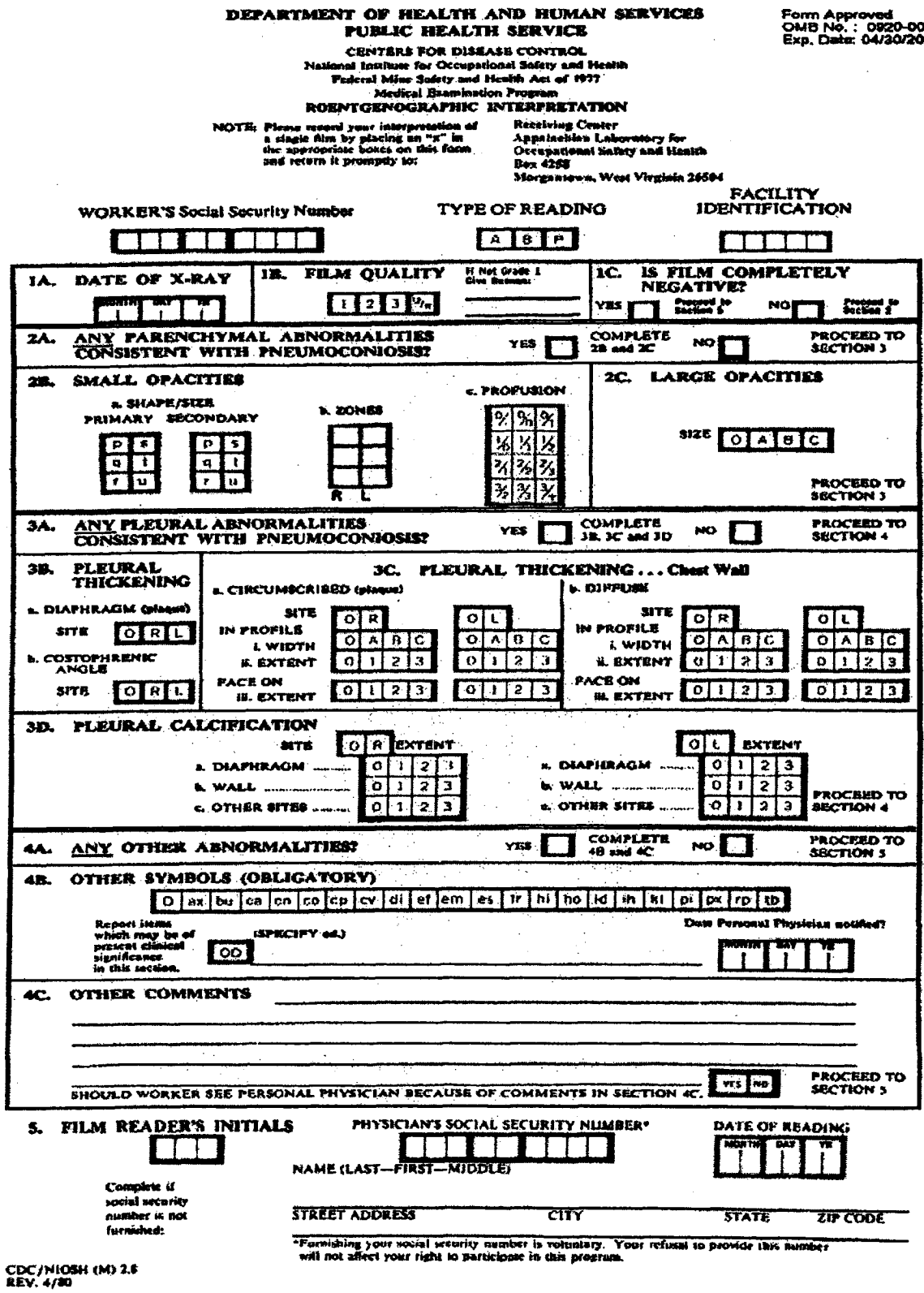
The ILO classification system is a guideline for readers of chest radiographs to record their findings related to pneumoconiosis. This includes small opacities associated with parenchymal changes, by shape and size of nodules, by their location in zones of each lung field, and by profusion using a 12-point scale from 0/− for completely normal lungs to 3/+ for advanced disease. The ILO system also requires readers to report the full gamut of changes in the lung from sources other than mineral dust retention. In its 1980 version, the ILO defined a reading of 1/0 as reflecting a viewer's opinion that the radiograph reveals some indications of pneumoconiotic changes. This level generally has been taken as "positive" evidence in the adjudication of compensation claims. The difference between minor categories (ie, 0/1 or 1/0) can be subtle and, in any given instance, a matter of the reader's judgment.

MATERIALS AND METHODS

The results presented in this article are based on a reading trial the authors conducted for a group of defense attorneys. The films in the study were obtained originally by plaintiffs' attorneys, read by initial readers of their choice, and, on the basis of the reader's conclusion, used by the attorneys to support litigation on behalf of the individual represented on the film. In accordance with legal procedures, the plaintiffs were required to submit both the film and the original reading for review by defendants, who are entitled to select their own experts. In effect, the group of consultant readers served as a cumulative expert. However, the consultants were not told the source of the films and were not told that the films had already been entered as evidence in litigation.

Seven groups of films and initial reports, totaling 551 cases, were made available to the authors from several legal sources. The authors were not given the names of the plaintiffs' law firms nor was demographic material provided about the individuals who were examined. As the groups of films were received, each case was assigned a study code. Some film jackets contained several sets of chest radiographs of the individual examined. In these instances, the most recent PA film that could be matched with an initial reading was selected for review by the consultant readers. The comparisons were between a single reading by 1 of 30 initial readers selected by plaintiffs and 6 independent readings of the same film by the consultants.

Six consultant radiologists and pulmonologists were invited to participate in the reading trial, and a seventh,



DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL
National Institute for Occupational Safety and Health
Federal Mine Safety and Health Act of 1977
Medical Examination Program
ROENTGENOGRAPHIC INTERPRETATION

Form Approved
OMB No. : 0920-0020
Exp. Date: 04/30/2001

NOTE: Please record your interpretation of a single film by placing an "X" in the appropriate boxes on this form and return it promptly to:

Receiving Center
Appalachian Laboratory for Occupational Safety and Health
Box 4258
Morgantown, West Virginia 26504

WORKER'S Social Security Number | TYPE OF READING (A B P) | FACILITY IDENTIFICATION

1A. DATE OF X-RAY (MONTH DAY YR)
1B. FILM QUALITY (1 2 3 9/R) — If Not Grade 1 Give Reasons:
1C. IS FILM COMPLETELY NEGATIVE? YES (Proceed to Section 5) NO (Proceed to Section 2)

2A. ANY PARENCHYMAL ABNORMALITIES CONSISTENT WITH PNEUMOCONIOSIS? YES (COMPLETE 2B and 2C) NO (PROCEED TO SECTION 3)

2B. SMALL OPACITIES
a. SHAPE/SIZE PRIMARY SECONDARY (p q r / s t u)
b. ZONES (R L)
c. PROFUSION (0/- 0/1 1/0 1/1 1/2 2/1 2/2 2/3 3/2 3/3 3/+)

2C. LARGE OPACITIES
SIZE (0 A B C)
PROCEED TO SECTION 3

3A. ANY PLEURAL ABNORMALITIES CONSISTENT WITH PNEUMOCONIOSIS? YES (COMPLETE 3B, 3C and 3D) NO (PROCEED TO SECTION 4)

3B. PLEURAL THICKENING
a. DIAPHRAGM (plaque) SITE (O R L)
b. COSTOPHRENIC ANGLE SITE (O R L)

3C. PLEURAL THICKENING . . . Chest Wall
a. CIRCUMSCRIBED (plaque)
b. DIFFUSE
SITE IN PROFILE (O R / O L)
i. WIDTH (0 A B C)
ii. EXTENT (0 1 2 3)
FACE ON iii. EXTENT (0 1 2 3)

3D. PLEURAL CALCIFICATION
SITE (O R) EXTENT (0 1 2 3)
a. DIAPHRAGM
b. WALL
c. OTHER SITES
SITE (O L) EXTENT (0 1 2 3)
a. DIAPHRAGM
b. WALL
c. OTHER SITES
PROCEED TO SECTION 4

4A. ANY OTHER ABNORMALITIES? YES (COMPLETE 4B and 4C) NO (PROCEED TO SECTION 5)

4B. OTHER SYMBOLS (OBLIGATORY)
(0 ax bu ca cn co cp cv di ef em es ir hi ho id ih kl pi px rp rb)
Report items which may be of personal clinical significance in this section. (SPECIFY od.) OD
Date Personal Physician notified? (MONTH DAY YR)

4C. OTHER COMMENTS
SHOULD WORKER SEE PERSONAL PHYSICIAN BECAUSE OF COMMENTS IN SECTION 4C. (YES NO)
PROCEED TO SECTION 5

5. FILM READER'S INITIALS
PHYSICIAN'S SOCIAL SECURITY NUMBER*
DATE OF READING (MONTH DAY YR)
NAME (LAST—FIRST—MIDDLE)
Complete if social security number is not furnished:
STREET ADDRESS CITY STATE ZIP CODE
*Furnishing your social security number is voluntary. Your refusal to provide this number will not affect your right to participate in this program.

CDC/NIOSH (M) 2.6
REV. 4/80

Figure 1. Interpretation form, National Institute for Occupational Safety and Health. ILO classification.

Table 1a
Technical Film Quality Grades Reported by Initial Readers

Initial Readers	Good		Acceptable		Poor		Unreadable		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
I-1	82	83.7	16	16.3	0	0	0	0	0	0	98	100.0
I-2	27	60.0	17	37.8	0	0.0	0	0	1	2.2	45	100.0
I-3	36	70.6	13	25.5	2	3.9	0	0	0	0	51	100.0
I-4	11	15.9	36	52.2	22	31.9	0	0	0	0	69	100.0
I-5	20	52.6	15	39.5	3	7.9	0	0	0	0	38	100.0
I-6	37	77.1	0	0.0	11	22.9	0	0	0	0	48	100.0
I-7	35	100.0	0	0.0	0	0.0	0	0	0	0	35	100.0
Other	58	68.2	20	23.5	7	8.2	0	0	0	0	85	100.0
Not stated	21	91.3	2	8.7	0	0.0	0	0	0	0	23	100.0
Total	327	66.5	119	24.2	45	9.1	0	0.0	1	0.2	492	100.0

Table 1b
Technical Film Quality Grades Reported by Consultant Readers

Consultant Readers	Good		Acceptable		Poor		Unreadable		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
C-1	75	15.2	228	46.3	117	23.8	72	14.6	0	0	492	100.0
C-2	22	4.5	320	65.0	129	26.2	20	4.1	1	.2	492	100.0
C-3	155	52.5	121	41.0	12	4.1	7	2.4	0	0	295	100.0
C-4	195	39.6	206	41.9	65	13.2	25	5.1	1	.2	492	100.0
C-5	239	48.6	235	47.8	13	2.6	4	0.8	1	.2	492	100.0
C-6	107	54.3	82	41.6	4	2.0	4	2.0	0	0	197	100.0
C-7	194	39.4	233	47.4	50	10.2	15	3.0	0	0	492	100.0
Total	987	33.4	1425	48.3	390	13.2	147	5.0	3	0.1	2952	100.0

who read 197 films, was asked to substitute when one of the original consultants could not continue. The readers were "blinded" as to the identities of study sponsors, the possible litigants, the previous or initial readers, and the individuals whose examinations they interpreted. They were asked to record their findings according to the ILO 80 classification system using the form devised by NIOSH (Fig. 1). They were not told the identities of the other readers and were not given the results of their individual or cumulative efforts. The consultant readers were compensated for their participation.

To ensure the "matching" of radiographs and reports, the worker's social security number and date of examination were recorded for each radiograph before the study code (the group letters and a consecutive number) was assigned. The identification information on the film was masked and the study code and social security number were recorded on each blank NIOSH form furnished to the consultant readers. The masking protected the anonymity of both those whose films were being read and the identity of initial readers that might have been noted on the flash card. The groups of films and matching forms were circulated to each of the six consultant readers for their independent review and recording of findings. A secure database was constructed to record and store the information received on each report form.

Using stringent criteria, the authors were able to identify 492 chest radiographs that matched an initial reader's report and were interpreted by six consultant readers. The initial readers' reports usually were recorded on a NIOSH form or were medical records that contained sufficient information to comply with the NIOSH form requirements. When the identification of the patient was verified on the films and the dates of x-ray examination and interpretation were confirmed, the case was accepted for inclusion in this report. Selected items were entered into a database, and the consultants' reading sheets were stored in a secure file. At the conclusion of the six consultants'

Table 1c
Agreement by Consultants With Initial Readers in Grading Chest X-Ray Film Quality

	Six Consultant Readers						
Initial Readers	Good	Acceptable	Poor	Unreadable	Not Stated	Total Readings	Percent Agreement
Good 327	**771**	922	220	46	3	1962	39.3
Acceptable 119	176	**387**	111	40	0	714	54.2
Poor 45	37	113	**59**	61	0	270	21.9
Unreadable 0	0	0	0	**0**	0	0	0
Not stated 1	3	3	0	0	**0**	6	0
Total 492	987	1425	390	147	3	2952	41.2

Table 1d
Assessment of Agreement Between Consultants and Initial Readers in Grading Chest X-ray Film Quality

Initial Readers' Film Grades	Number of Good Ratings by Six Consultant Readers							Total Films
	0	1	2	3	4	5	6	
Good	48	60	64	71	45	31	5	324
Other	59	48	30	10	11	5	1	164
Total	107	108	94	81	56	36	6	488

readings, the films were stripped of the study code and returned to their sources.

Statistical Methods

The basic data in this report were recorded on the NIOSH form by each consultant reader and by most of the initial readers. The instructions on the form and the ILO guidelines influence both the interpretation of the chest x-ray films by the readers and the accuracy and completeness of their reports. Examples of this influence are noted in Tables 1a and 1b and Tables 2a and 2b, showing marked differences between the initial readers and the consultants related to the grading of film quality and the answers to the question "Is film completely negative?" The implications of an "unreadable" grade for film quality and a "yes" answer to "Is film completely negative?" are similar in that the instructions indicate that no further substantive entries should be made on the form by the readers. A standard approach was used to accommodate this "skip" pattern. For entries in which a positive response was recorded for "Is film completely negative?" the remaining entries about abnormalities on the film were coded to reflect the absence of abnormalities. When the "film quality" was graded as "unreadable," the remaining entries were coded as "missing."

It also should be noted that on Tables 1a and 2a, each of the seven initial readers who read 35 or more films are shown separately. They accounted for a total of 384 reports. The 108 remaining reports were provided by 23 initial readers, who were grouped as "other" or "not stated."

Agreement tables are provided comparing data between initial and consultant readers for film quality grades, film completely negative, parenchymal abnormalities, small opacities profusion, and pleural abnormalities. To assess whether or not the initial readers and the consultants showed tendencies toward different interpretations of these items, odds ratios and kappa statistics were estimated (12,13). Because the study has a 1:M matching design, Mantel-Haenszel chi-square statistics were estimated (14). Ninety-five percent confidence intervals were used to assess significance of odds ratios. To assess agreement among the six consultants, and then with the inclusion of the initial reader's report, a coefficient of agreement was computed and evaluated in accordance with Landis and Koch's measurement of observer agreement (15). Kappa statistics were also used to compare the median value for the six consultants to assess agreement with the initial readers' ratings. For films with a median rating that was not integer-valued (eg, some median rat-

Table 2a
Is Film Completely Negative? Recorded by Initial Readers

Initial Readers	Yes		No		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%
I-1	0	0.0	98	100.0	0	0.0	98	100.0
I-2	0	0.0	45	100.0	0	0.0	45	100.0
I-3	0	0.0	51	100.0	0	0.0	51	100.0
I-4	0	0.0	69	100.0	0	0.0	69	100.0
I-5	0	0.0	38	100.0	0	0.0	38	100.0
I-6	0	0.0	48	100.0	0	0.0	48	100.0
I-7	0	0.0	35	100.0	0	0.0	35	100.0
Other	0	0.0	85	100.0	0	0.0	85	100.0
Not stated	0	0.0	23	100.0	0	0.0	23	100.0
Total	0	0.0	492	100.0	0	0.0	492	100.0

Table 2b
Is Film Completely Negative? Recorded by Consultant Readers

Consultant Readers	Yes		No		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%
C-1	130	26.4	290	58.9	72	14.6	492	100.0
C-2	124	25.2	348	70.7	20	4.1	492	100.0
C-3	152	51.5	136	46.1	7	2.4	295	100.0
C-4	218	44.3	248	50.4	26	5.3	492	100.0
C-5	159	32.3	329	66.9	4	.8	492	100.0
C-6	140	71.1	53	26.9	4	2.0	197	100.0
C-7	197	40.0	280	56.9	15	3.1	492	100.0
Total	1120	37.9	1684	57.1	148	5.0	2952	100.0

ings were 1.5, meaning that three raters gave a rating of 1, whereas the other three rated it a 2), the median was randomly rounded to either the upper or lower integer.

RESULTS

The results are based on analyses of a dataset of 492 film readings by initial readers and 2,952 interpretations by the consultants. After applying the first matching process described previously, the data from the 551 initial readers' reports were used to produce preliminary tables and analyses. When it was decided to produce additional tables to perform "Agreement" analyses, more stringent criteria were applied to the matching process that reduced the number of films to 492.

Summary and Agreement Tables

The "matched" dataset Tables 1a and 1b show important differences in technical film quality grades assigned by the consultants and initial readers. Based on the data set, "good" film quality was recorded for 66.5% of the chest x-ray examinations by the initial readers, compared with 33.4% by the consultants. The grade of "acceptable" quality was reported for 24.2% of the films by initial readers and 48.3% by the consultants. The initial readers recorded "poor" quality for 9.1%, whereas the consultants assigned "poor" quality to 13.2% of the films. There were no "unreadable" film quality grades reported by an initial reader, compared with 147 or 5% by the consultants. Individually, the initial readers' range of "good" grades was 15.9–100% and 0–52.2% for "acceptable." The consultants' "good" quality grades ranged from 4.5% to 54.3% and from 41.0 to 65.0% for "acceptable."

The data in Table 1c are based on the 492 cases that were subjected to the stringent matching process. Here, each of the consultant ratings of film quality was compared with the specified rating by the initial readers; "agreement" is indicated in the table by the bold num-

Table 2c
Agreement by Consultants With Initial Readers' Interpretations for Answers to "Is Film Completely Negative?"

Initial Readers	Six Consultant Readers			Total	Percent Agreement
	Yes	No	Not Stated		
Yes 0	0	0	0	0	0
No 492	1120	1684	148	2952	60.1
Not stated 0	0	0	0	0	0
Total 492	1120	1684	148	2952	60.1

Table 2d
Assessment of Agreement Between Consultants and Initial Readers' Interpretations for Answers to "Is Film Completely Negative?"

Initial Readers' Answers	Number of "Yes's" by Six Consultant Readers							Total Films
	0	1	2	3	4	5	6	
Yes	1	0	0	0	0	0	0	1
No	116	61	39	43	43	57	39	398
Total	117	61	39	43	43	57	39	399

bers. For the 327 films rated "good" by the initial readers, the consultants rated the quality "good" on 771 reports, an agreement percentage of 39.3. The consultants agreed with the initial readers' "acceptable" ratings on 387 reports, a percentage of 54.2. Agreement on "poor" ratings by the consultants was 59 reports or 21.9%. For all of the film quality categories, consultant agreement with initial readers was 41.2%. As indicated, none of the initial readers rated any of the films "unreadable."

For the assessment in Table 1d, the data for film quality were dichotomized as "good" or "other" (ie, "acceptable," "poor," and "unreadable" were grouped as "other"). One "not stated" rating by an initial reader and three films with "not stated" ratings by the consultants were excluded for a total of 488 films. In this table, the rows are indexed by whether or not the initial reader rated the quality of a film as "good"; the columns are indexed by how many of the consultant readers gave the film a "good" rating. For example, there were 48 films that received a "good" rating by the initial reader and a different rating by each of the six consultants. Similarly, there were 11 films that did not receive a "good" rating by the initial reader, but were rated "good" by 4 of the consultant readers.

The odds ratio associated with the data in Table 1d is 5.53 with a 95% confidence interval of 4.36–7.02 ($P <$ .0001). This indicates a probability of less than 1 in 10,000 that the differences noted between initial and consultant readers are due to chance alone. The odds that an initial reader will rate film quality as "good" are 5.53 times the odds that a consultant reader will rate the same film as "good" quality. These findings show that an initial reader is much more likely to conclude that a chest x-ray is "good" than is a consultant reader.

The kappa statistic assessing agreement among the six consultant readers was .10 with a 95% confidence interval of .08–.12. When the initial reader is included in the analysis, the kappa statistic for the seven readers drops to .09 with a 95% confidence interval of .07–.10. This implies that the agreement for film quality ratings among readers is "slight." The kappa statistic comparing the median consultant scores with the initial readers' scores was estimated to be .20 with a 95% confidence interval of .14–.26. This suggests that the agreement for film quality between consultants and initial readers is also "slight."

In Tables 2a and 2b, consultants answered the question "Is Film Completely Negative" "yes" on 37.9% of their interpretive reports, compared with 0 reports by initial readers. For individual consultants, the range of "completely negative" reports was 25.2–71.1%. Consultants answered "no" to this question on 57.1% of their reports, compared with 100.0% for initial readers. It should be noted that the NIOSH form instructs readers who indicate that a chest x-ray is "completely negative" to make no additional substantive entries on the form. The 1,120 completely negative reports noted by consultants were

Table 3a
Any Parenchymal Abnormalities Consistent With Pneumoconiosis

Parenchymal Abnormalities	Initial Readers		Consultant Readers	
	Count	%	Count	%
Yes	478	97.2	182	6.2
No	14	2.8	2610	88.4
Not stated	0	0	160	5.4
Total	492	100	2952	100

considered as "no" answers in Tables 3a (parenchymal) and 3b (pleural abnormalities).

In the matched set of 492 cases (Table 2c), initial readers recorded "no" on each report. The consultants agreed with this answer on 1,684 report forms, an agreement percentage of 60.1. It should be noted that the 148 readings shown as "not stated" for the consultants includes 147 graded as "unreadable" for film quality. The 148 "not stated" ratings were excluded from the total of 2,952 in calculating the percent agreement.

For this assessment (Table 2d), the data for "Is Film Completely Negative?" are dichotomized as "yes" or "no." Films stated as having "unreadable" quality and those with missing responses by at least one reader were excluded, leaving 399 for analysis. In this table, the rows are indexed by whether or not the initial reader rated the film as completely negative; the columns are indexed by how many of the consultant readers gave the film a "yes" to being completely negative. For example, of the 398 "no" answers by the initial readers, there were 116 films in which all six consultants agreed with the answers. For 39 films with "no" by initial readers, all six consultants disagreed.

The odds ratio associated with the data in Table 2d is .0063 with a 95% confidence interval of .0020–.020 ($P <$ .0001). This indicates a probability of less than 1 in 10,000 that the differences noted between initial and consultant readers are due to chance alone. These findings show that an initial reader is much more likely to conclude that a chest x-ray is not completely negative than is a consultant reader. Specifically, a consultant reader has 159 times the odds of concluding that a film is completely negative than an initial reader.

The kappa statistic assessing agreement among the six consultant readers was .43 (95% CI .40–.46), and, after including the initial readers, the kappa statistic for the seven readers drops to .31 (95% CI .29–.34). This implies that the agreement for "Is Film Completely Negative?" ratings among consultant readers is "moderate"; when initial readers' ratings are included, the agreement declines to "fair." The kappa statistic comparing the median consultant scores to the initial readers' scores was estimated to be −.005 (95% CI −.017 to .0072), suggesting practically no agreement between consultants and initial readers.

Initial readers answered the question in Table 3a "yes" in 97.2% of their interpretive reports on the chest x-ray examinations and "no" in 2.8%. Individually, the initial readers' range of positive reports was 84–100%. Consultants recorded "yes" in 6.2% and "no" in 88.4% of their reports. The range of positive reports by individual consultants was 3.4–11.5%. As a result of the number of "unreadable" film quality ratings by the consultants, "not stated" was indicated in 5.4% of their reports compared with none by initial readers.

For the assessment in Table 3b, the data for parenchymal abnormalities were reported as "yes" or "no," indicating their presence in the chest x-ray. Eighty films with "unreadable" film quality ratings and 12 films in which initial or consultant readers did not respond to the question were excluded to yield a total of 400 films. In this table, the rows are indexed by whether or not the initial reader said there were parenchymal abnormalities in the film; the columns are indexed by how many of the con-

Table 3b
Assessment of Agreement Between Consultants and Initial Readers' Interpretations of Parenchymal Abnormalities

Initial Readers' Film Count	Number of "Yes's" by Six Consultant Readers							Total Films
	0	1	2	3	4	5	6	
Yes	315	39	18	9	1	4	2	388
No	11	1	0	0	0	0	0	12
Total	326	40	18	9	1	4	2	400

Table 4a
Parenchymal Abnormalities—Small Opacities Profusion

Small Opacities Profusion Category	Initial Readers		Consultant Readers	
	Count	%	Count	%
0/0	0	0	1	0.0
0/1	5	1.0	48	1.6
1/0	326	66.3	52	1.8
1/1	120	24.4	50	1.7
1/2	14	2.8	11	.4
2/1	6	1.2	5	.2
2/2	3	.6	4	.1
2/3	1	.2	5	.2
3/2	0	0.0	1	0.0
3/3	1	.2	5	.2
3/+	1	.2	0	0.0
Not stated	15	3.0	2770	93.8
Total	492	100.0	2952	100.0

sultant readers said there were parenchymal abnormalities in the same film. For example, there were nine films that received a "yes" rating by the initial reader and the same rating by three of the six consultants.

The odds ratio associated with the data in Table 3b is 2184.0 with a 95% confidence interval of 282.9–16,859.1 ($P < .0001$). These findings indicate the odds that an initial reader will answer "yes" to the question "Any Parenchymal Abnormalities Consistent With Pneumoconiosis" are 2184.0 times the odds that a consultant reader will answer "yes" to the question for the same film.

The kappa statistic assessing agreement among the six consultant readers was .31 with a 95% confidence interval of .28–.33. When the initial readers are included in the analysis, the kappa statistic for the seven readers drops to −.027 with a 95% confidence interval of −.049 to −.0055. This implies that the agreement for parenchymal abnormalities among consultant readers is "fair" and becomes "poor" when initial readers' interpretations are included. The kappa statistic comparing the median consultant scores to the initial readers' scores was estimated to be .0019 with a 95% confidence interval of −.0037 to .0075, suggesting complete lack of agreement.

In the matched dataset (Table 4a), the initial readers recorded profusion categories of 1/0 or higher in 472, or 95.9% of their 492 interpretations. Of these, 326, or 66.3%, were shown as "category 1/0" and 120 or 24.4% as "category 1/1." The six consultant readers recorded profusion categories of 1/0 or higher, in the matched data set, on 133 or 4.5% of their 2,952 reports. Of these, 48, or 1.6%, were shown as "category 0/1"; 52, or 1.8%, were "category 1/0"; and 50, or 1.7%, were "category 1/1." The "not stated" category includes 15 films interpreted by initial readers and 2,770 readings by consultants that were not categorized for small opacities profusion. Of these, 14 by initial readers and 2,610 by consultants were associated with reports of "no" parenchymal abnormalities, as indicated in Table 3a. The remaining "not stated" cases are related to "unreadable" films and lack of entries indicated in previous tables.

For the assessment in Table 4b, the data for small opacities profusion were dichotomized in two categories: "less than 1/0" and "1/0 and greater." Small opacities that are "1/0 or greater" may be considered for legal compensation. Eighty films with "unreadable" film quality ratings and 18 films in which initial or consultant readers did not respond to the question were excluded to yield a total of 394 films. In this table, the rows are indexed by whether the initial reader rated the small opacities as "less than 1/0" or "1/0 and greater"; the columns are indexed by how many of the six consultant readers gave the small opacities a rating of "1/0 or greater." This table shows that there were 322 films that received a "1/0 or greater" rating by the initial reader and a different rating by each of the six consultants. There was 1 of the 394 films rated as "1/0 or greater" by an initial reader in which all six consultants agreed.

Table 4b
Assessment of Agreement Between Consultants and Initial Readers' Ratings of Small Opacities Profusion

Initial Readers' Ratings	"1/0 or greater" Ratings by Six Consultant Readers							Total Films
	0	1	2	3	4	5	6	
Less than 1/0	14	2	0	0	0	0	0	16
1/0 or greater	322	37	9	5	1	3	1	378
Total	336	39	9	5	1	3	1	394

Table 5a
Any Pleural Abnormalities Consistent With Pneumoconiosis?

Pleural Abnormalities	Initial Readers		Consultant Readers	
	Count	%	Count	%
Yes	130	26.4	224	7.6
No	358	72.8	2568	87.0
Not stated	4	0.8	160	5.4
Total	492	100.0	2952	100.0

The odds ratio associated with the data in Table 4b is 2,227.0 with a 95% confidence interval of 294.5–16,840.1 ($P < .0001$). These findings indicate the odds that an initial reader will rate the small opacities as "1/0 or greater" are 2,227.0 times the odds that a consultant reader will agree on the same rating. An initial reader is much more likely to conclude that the small opacities is "1/0 or greater" than is a consultant reader.

The kappa statistic assessing agreement among the six consultant readers was .19 with a 95% confidence interval of .18–.21. When the initial readers are included in the analysis, the kappa statistic for the seven readers drops to −.037 with a 95% confidence interval of −.053 to −.021. This implies that the agreement for small opacities profusion ratings among consultant readers is "slight," and when initial readers' ratings are included, the agreement becomes "poor." The kappa statistic comparing the median consultant scores with the initial readers' scores was estimated to be .025 with a 95% confidence interval of .017–.032. This suggests that there is essentially no agreement between consultants and initial readers in small opacities profusion ratings.

Based on the matched data set, initial readers answered the question in Table 5a "yes" on 130 reports or 26.4% of their 492 interpretations of chest x-ray examinations, and "no" on 358 or 72.8% of their reports. The consultant readers recorded "yes" on 224 or 7.6% of their reports, and "no" on 2,568 or 87.0% of their reports. For individual initial readers, the range of positive responses was 7.9–92.2% and 7.8–91.9% for negative answers. The positive response range for consultants was 6.1–12.9%.

For the assessment in Table 5b, the data for pleural abnormalities were reported as "yes" or "no," indicating their presence in the chest x-ray. Eighty films with "unreadable" film quality ratings and 21 films in which initial or consultant readers did not respond to the question were excluded to yield a total of 391 films. In this table, the rows are indexed by whether or not the initial reader said there were pleural abnormalities in the chest x-ray; the columns are indexed by how many of the consultant readers said there were pleural abnormalities in the same film. For example, there were 256 films that initial readers reported having "no" pleural abnormalities, for which none of the consultants recorded "yes."

The odds ratio associated with the data in Table 5b is 9.2 with a 95% confidence interval of 6.2–3.6 ($P < .0001$). These findings indicate the odds that an initial reader will answer "yes" to the question "Any Pleural Abnormalities Consistent with Pneumoconiosis?" are 9.2 times the odds that a consultant reader will answer "yes" to the question regarding the same film. An initial reader is much more likely to conclude that a chest x-ray has pleural abnormalities than a consultant reader.

The kappa statistic assessing agreement among the six consultant readers was .49 with a 95% confidence interval of .47–.52. When the initial readers are included in the analysis, the kappa statistic for the seven readers drops to .36 with a 95% confidence interval of .34–.39. This implies that the agreement for film quality ratings among consultant readers is "moderate," and when initial readers are included the agreement declines to "fair." The kappa statistic comparing the median consultant scores to the initial readers' scores was estimated to be .25 with a 95% confidence interval of .17–.32. This suggests that the agreement between consultants and initial readers for the presence of pleural abnormalities is "fair."

Table 5b
Assessment of Agreement Between Consultants and Initial Readers' Interpretations of Pleural Abnormalities

Initial Readers' Interpretations	Number of "Yes" Answers by Six Consultant Readers							Total Films
	0	1	2	3	4	5	6	
Yes	64	7	8	12	3	2	7	103
No	256	23	2	2	4	1	0	288
Total	320	30	10	14	7	3	7	391

Table 6
Parenchymal and Pleural Abnormalities Consistent With Pneumoconiosis

Parenchymal and Pleural Abnormalities	Initial Readers		Consultant Readers	
	Count	%	Count	%
Yes	120	24.4	27	.9
No	354	72.0	153	5.2
Other	18	3.7	2772	93.9
Total	492	100.0	2952	100.0

Based on the matched dataset (Table 6), 120 or 24.4% of the reports by initial readers noted both parenchymal and pleural abnormalities on a chest x-ray examination. For the consultant readers, 27 or .9% of their reports showed both types of abnormalities on their interpretations of the radiographs.

In the matched data set (Tables 7a and 7b), 174 or 35.4% of the initial readers indicated findings of clinical importance other than parenchymal and pleural abnormalities that were consistent with pneumoconiosis. The consultant readers answered this question "yes" on 1,552 or 52.6% of their reports. The range of positive responses to this question by individual initial readers was 11.4–66.7%. For the individual consultants, the range of positive responses was 12.7–65.9%.

To facilitate the recording of other abnormalities of clinical importance, the ILO 80 guide provides "symbols" that are on the NIOSH form as boxes to be checked when appropriate (Table 8). The distribution of the 149 symbols checked by initial readers to indicate other abnormalities is shown in the table and indicates that the most frequently reported findings were pleural thickening in the interlobular fissure or mediastinum, 28.9%; abnormality of cardiac size or shape, 22.1%; and definite emphysema, 10.1%. For the consultants, the distribution of the 904 symbols checked shows the most frequent as abnormality of cardiac size or shape, 38.1%; definite emphysema, 20.6%; and bulla(e), 7.6%.

DISCUSSION

This study includes a relatively large number of radiographs and interpretations. They were obtained as noted in a manner that did not allow any consideration of demographic factors. All of the individuals examined had in common an asserted work history involving asbestos and a decision by an attorney that a legal claim of damage from occupational exposure to asbestos had a potential of obtaining compensation. This potential was based on the initial B reading, plus any other available medical evidence.

In terms of our objective of comparing or contrasting initial interpretations by plaintiffs' readers with the conclusions of the consultants, demographic factors were of interest but did not govern. The authors had no basis for determining how this group of workers represented the universe of asbestos claimants. The "universe" is a constantly shifting group as new cases are filed frequently. It has been more than 30 years since the US government imposed restraints on asbestos and many companies that had manufactured or used asbestos products shifted to other, presumably safer, materials. However, as noted in

Table 7a
Any Other Abnormalities? Recorded by Initial Readers

Initial Readers	Yes		No		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%
I-1	42	36.5	71	61.7	2	1.7	115	100.0
I-2	30	66.7	15	33.3	0	0.0	45	100.0
I-3	9	17.6	41	80.4	1	2.0	51	100.0
I-4	22	31.9	47	68.1	0	0.0	69	100.0
I-5	11	28.9	25	65.8	2	5.3	38	100.0
I-6	28	58.3	19	39.6	1	2.1	48	100.0
I-7	4	11.4	30	85.7	1	2.9	35	100.0
Other	21	30.9	44	64.7	3	4.4	68	100.0
Not stated	7	30.4	16	69.6	0	0.0	23	100.0
Total	174	35.4	308	62.6	10	2.0	492	100.0

Table 7b
Any Other Abnormalities? Recorded by Consultant Readers

Consultant Readers	Yes		No		Not Stated		Total	
	Count	%	Count	%	Count	%	Count	%
C-1	279	56.7	11	2.2	202	41.1	492	100.0
C-2	324	65.9	5	1.0	163	33.1	492	100.0
C-3	131	44.4	3	1.0	161	54.6	295	100.0
C-4	224	45.5	22	4.5	246	50.0	492	100.0
C-5	304	61.8	24	4.9	164	33.3	492	100.0
C-6	25	12.7	26	13.2	146	74.1	197	100.0
C-7	265	53.9	15	3.0	212	43.1	492	100.0
Total	1552	52.6	106	3.6	1294	43.8	2952	100.0

Table 8
Symbols for Other Abnormalities Recorded by Initial and Consultant Readers

Symbol	Initial Readers		Consultant Readers	
	Total	%	Total	%
ax: Coalescence of small pneumoconiotic opacities	1	0.7	0	0
bu: Bulla(e)	8	5.4	69	7.6
ca: Cancer of lung or pleura	14	9.4	29	3.2
cn: Calcification in small pneumoconiotic opacities	1	0.7	0	0.0
co: Abnormality of cardiac size or shape	33	22.1	344	38.1
di: Marked distortion of the intrathoracic organs	1	0.7	28	3.1
ef: Effusion	4	2.7	28	3.1
em: Definite emphysema	15	10.1	186	20.6
fr: Fractured rib(s)	7	4.7	65	7.2
hi: Enlargement of hilar or mediastinal lymph nodes	3	2.0	25	2.8
ho: Honeycomb lung	2	1.3	13	1.4
id: Ill-defined diaphragm	8	5.4	13	1.4
ih: Ill-defined heart outline	7	4.7	23	2.5
kl: Septal (Kerley) lines	1	0.7	6	0.7
pi: Pleural thickening in the interlobular fissure or mediastinum	43	28.9	63	7.0
tb: Tuberculosis	1	0.7	12	1.3
Total	149	100.0	904	100.0

the proceedings of the July 2003 US Senate Judiciary Committee hearing on S 1125, the Fairness in Asbestos Injury Resolution Act of 2003, the volume of new claims has increased significantly in the last few years (16).

Additional studies incorporating scientific sample selection and the use of demographic factors would be desirable in developing more objective measures for adjudication of individual claims. To develop reliable estimates of the incidence and prevalence of asbestos-related changes in a specified group of workers who may have been occupationally exposed to asbestos, demographic data are needed. These include such characteristics as age, length of employment, extent of exposure, and smoking history. Despite these limitations, the results of this study can contribute to the design of subsequent efforts to obtain reliable information about asbestosis, including the optimum number of consultant readers needed to provide sufficient datapoints to support reasonable analysis and conclusions.

Interobserver variation in the interpretation of medical images has been a concern since the earliest days of radiology. In classic studies in the 1950s, Yerushalmy et al. and Garland et al. (17,18) indicated a 30% variance in findings among physicians interpreting radiographs of a selected batch of proven cases. They also demonstrated a 10% intraobserver variance when the same radiologist reinterpreted a set of radiographs after a time delay.

Early in the national coal worker chest surveillance program in 1973, Felson et al. and Trout et al. (19,20) commented on findings from the multiple reading system devised by NIOSH for review of chest radiographs entered into its programs. NIOSH still uses a multiple reader scheme involving two or three B readers for its coal miner programs. Ducatman et al., (21) in their study of the range of B reader interpretations that included 23 observers and a total of 105,000 readings, concluded that individual diagnoses, legal decisions, and population assessments ought to rely on multiple readings of chest radiographs.

In addition to the present study, the authors reviewed the world literature on chest x-ray studies of lung changes related to mineral dust retention in various worker populations (22–37). These reports discussed studies in several countries of worker populations exposed to asbestos and other mineral dusts. Most of the studies used the ILO 80 system as a means of recording their findings. Allowing for variations in technique, the studies cited involved interpretation of radiographs by more than one expert reader with aggregate results reported. In no instance did the conclusions agree with the high level of positivity (1/0 or higher on the ILO scale for small opacities) reported by the initial readers in the current study. The proportion of parenchymal abnormality—small opacity profusion ratings of 1/0 and higher—reported by the initial readers were significantly higher than those recorded by the consultant readers. The findings in the literature, which apply to asbestos worker groups and those exposed to other mineral dusts, support the results of the consultants' interpretations in this study rather than the high levels of positivity reported by the initial readers.

CONCLUSION

Reinterpretation by six independent consultants of chest radiographs read initially by B readers selected by plaintiffs' counsel failed to confirm the conclusions of the initial readers. Whereas the initial readers interpreted 95.9% of the chest x-rays as positive for parenchymal abnormalities—small opacities profusion category 1/0 or higher (ILO 80)—the consultants interpreted the same set of cases as positive in only 4.5%. Comparisons of other pertinent data from the reports showed similar significant differences between the interpretations of the initial readers and those of the independent consultants.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of five individuals whose participation made the survey on which this paper is based possible: Diana Linton, Diane Cook, Edna Reffit, Michelle Fellers, and Anand Narayan. Their help was graciously given and gratefully received.

REFERENCES

1. ILO 1980 Classification of radiographs of the pneumoconioses. Geneva, Switzerland: Occupational Safety and Health Service, International Labor Office, publication 22, revised.
2. Specifications for medical examinations of underground coal miners: Chest roentgenographic examinations. 35 FR 42, chapter 1, subchapter C, 19 August 1970.
3. Morgan RH. Proficiency examination of physicians classifying pneumoconiosis chest films. AJR Am J Roentgenol 1979; 132:803–808.
4. 29 CFR, subtitle B, chapter XVII, part 1910, appendix E, 1999.
5. Hitt G. Asbestos makers litigants: uneasy victims. Wall Street Journal, 28 May 2002, p. A4; Sidel R, Warren S. Liabilities may trail asbestos spin offs. 29 May 2002, p. A1.
6. Samuel RJ. Asbestos fraud. Washington Post. 20 November 2002; p. A23.
7. Selikoff IJ, Churg J, Hammond EC. Asbestos exposure and neoplasia. JAMA 1964; 188:22–26.
8. Selikoff IJ, Churg J, Hammond EC. The occurrence of asbestosis among insulation workers in the US. Ann NY Acad Sci 1965; 132:139–155.
9. Asbestos related diseases—clinical, epidemiologic, pathologic and radiologic characteristics and manifestations. The Asbestos Working Group, American College of Radiology Task Force on Pneumoconiosis. Chicago: ACR; 1982.
10. American Thoracic Society. The diagnosis of nonmalignant diseases related to asbestosis. Am Rev Resp Dis 1986; 134:363–368.
11. Jones NR. Draft report of Commission on Asbestos Litigation, American Bar Association, Chicago, IL. February 2003.
12. Kundel H, Polansky M. Measurement of observer agreement. Radiology 2003; 228:303–308.
13. Sistrom CL, Garvan CW. Proportions, odds and risk. Radiology 2004; 230:12–19.
14. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22:719–748.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.
16. US Senate Committee on the Judiciary. Report 108–118, on S 1125, the Fairness in Asbestos Injury Resolution Act of 2003; 30 July 2003. p. 75.
17. Yerushalmy J, Harkness JT, Cope JH, et al. Role of dual reading in mass radiography. Am Rev Tuberc 1950; 61:443–464.
18. Garland LH, Miller ER, Zwerling HB, et al. Studies on the value of serial films in estimating the progress of pulmonary disease. Radiology 1952; 58:161–175.
19. Felson B, Morgan WKC, Bristol LJ, et al. Observations on the results of multiple readings of chest films in coal miners' pneumoconiosis. Radiology 1973; 109:19–23.
20. Trout ED, Jacobson G, Moore RT, et al. Analysis of the rejection rate of chest radiographs obtained during the coal mine "black lung" program. Radiology 1973; 109:25–27.
21. Ducatman AM, Yang WN, Forman SA. 'B-readers' and asbestos medical surveillance, radiographic interpretation of asbestosis. J Occup Med 1988; 30:644–647.
22. Maehle WM, Muir D, Chan JC, et al. The Canadian pneumoconiosis reading panel study. Proceedings, 7th International Conference on the Pneumoconioses, Pittsburgh, PA: August 1989. pp. 201–204.
23. Delclos GL, Wilson RK, Bradley BL. Interobserver variability using the ILO (1980) classification in subjects referred for compensation evalua-

tion proceedings. Proceedings, 7th International Conference on the Pneumoconioses, Pittsburgh, PA: August 1989. pp 960–964.
24. Reger RB, Cole WS, Sargent EN, et al. Cases of alleged asbestos-related disease, a radiologic re-evaluation. J Occup Med 1990; 32:1088–1090.
25. Koskinen K, Rinne JP, Zitting A, et al. Screening for asbestos-induced diseases in Finland. Am J Ind Med 1996; 30:241–251.
26. Garcia-Closas M, Christiani DC. Asbestos-related diseases in construction carpenters. Am J Ind Med 1995; 27:115–122.
27. Welsh LS, Michaels D, Zoloth SR. The national sheet metal examination group (1994). The national sheet metal worker asbestos disease screening program: radiological findings. Am J Ind Med 1994; 25:635–648.
28. Anton-Culver H, Culver BD, Kurosaki T. An epidemiological study of asbestos-related changes to identify work areas of high risk in a shipyard population. Appl Ind Hyg 1989; 4:110–118.
29. Merewether ERA, Price CW. Report on the effects of asbestos dust on the lungs and dust suppression in the asbestos industry, I—Occurrence of pulmonary fibrosis and other pulmonary affectations in asbestos workers. HM Stationery Office, London, UK; 1930.
30. Meurman LO, Kiviluoto R, Kakama M. Mortality and morbidity among the working population of anthophyllite asbestos miners in Finland. Br J Ind Med 1974; 31:105–112.
31. Donnelly J. Pulmonary asbestosis: Incidence and prognosis. Ind Hyg 1936; 18:222–228.
32. Shull JR. Asbestosis: A roentgenologic review of 71 cases. Radiology 1936; 27:279–292.
33. Kagamimori S, Matsusbara I, Sokejima S, et al. Studies on changes in categories for pneumoconiosis x-ray classification in Japanese workers with occupational exposure to mineral dusts. Proceedings of 9th International Conference on Occupational Respiratory Diseases (ICORD). Kyoto, Japan. Amsterdam, the Netherlands; Elsevier, 1997. p. 166–169.
34. Murai Y, Kitagawa M. Autopsy cases with asbestosis in Japan. Proceedings of 9th International Conference on Occupational Respiratory Diseases (ICORD). Kyoto, Japan. Amsterdam, the Netherlands; Elsevier, 1997. p. 242–245.
35. Tossavainen A, Huskonen MS, Koskinen K, et al. Radiographic abnormalities among Finnish construction, shipyard and asbestos industry workers. Proceedings of 9th International Conference on Occupational Respiratory Diseases (ICORD). Kyoto, Japan. Amsterdam, the Netherlands; Elsevier, 1997. p. 290–292.
36. Izmerov NF, Elovskaya LT, Turmistrova TV, et al. Roentgenographic and hygienic comparison of fibrosis induced by chrysotile-asbestos. Proceedings of 9th International Conference on Occupational Respiratory Diseases (ICORD). Kyoto, Japan. Amsterdam, the Netherlands; Elsevier, 1997. p. 334–336.
37. Hisanaga N, Shibata E, Sun J, et al. Pleural plaques and irregular opacities on chest radiographs among construction workers. Proceedings of 9th International Conference on Occupational Respiratory Diseases (ICORD). Kyoto, Japan. Amsterdam, the Netherlands; Elsevier, 1997. p. 286–289.